Journal of Structural and Functional Genomics **2:** 65–73, 2003.
© 2003 *Kluwer Academic Publishers. Printed in the Netherlands.*

Are all fishes ancient polyploids?

Yves Van de Peer[1]*, John S. Taylor[2†] & Axel Meyer[2]
[1] *Department of Plant Systems Biology, Vlaams Interuniversitair Instituut voor Biotechnologie (VIB), Ghent University, K.L. Ledeganckstraat 35, B−9000 Gent, Belgium*; [2] *Department of Biology, University of Konstanz, D−78457 Konstanz, Germany; † present address: Department of Biology, PO Box 3020, University of Victoria, Victoria, Canada, V8W 3N5*
* *Author for correspondence: E-mail: yves.vandepeer@gengenp.rug.ac.be*

Received 26.03.2002; accepted in final form 29.08.2002

Key words: genome duplication, gene evolution, subfunctionalization

Abstract

Euteleost fishes seem to have more copies of many genes than their tetrapod relatives. Three different mechanisms could explain the origin of these 'extra' fish genes. The duplicates may have been produced during a fish-specific genome duplication event. A second explanation is an increased rate of independent gene duplications in fish. A third possibility is that after gene or genome duplication events in the common ancestor of fish and tetrapods, the latter lost more genes. These three hypotheses have been tested by phylogenetic tree reconstruction. Phylogenetic analyses of sequences from human, mouse, chicken, frog (*Xenopus laevis*), zebrafish (*Danio rerio*) and pufferfish (*Takifugu rubripes*) suggest that ray-finned fishes are likely to have undergone a whole genome duplication event between 200 and 450 million years ago. We also comment here on the evolutionary consequences of this ancient genome duplication.

Introduction

Several authors have presupposed that major evolutionary transitions in biology have required the genetic raw material provided by gene, chromosome, and/or entire genome duplications (Ohno, 1970; Sidow, 1996; Spring, 1997; Holland, 1999; Lundin, 1999; Patel and Prince, 2000). Already about 30 years ago, Ohno (1970) presented comparative data on genome size and chromosome numbers to support his hypothesis that one or more genome duplications occurred during the evolution of vertebrates and made their diversification possible. Ohno hypothesized that big leaps in evolution – such as the transition from an invertebrate to a vertebrate – required the creation of new gene loci with previously non-existent functions and emphasized genome duplication via tetraploidy as the mechanism for the production of such new genes. Gene number comparisons do provide support for large-scale gene or genome duplication events in the vertebrate lineage. Spring (1997) uncovered an average of three homologous genes in humans for each of 52 genes of *Drosophila* and proposed that the additional human genes were produced during two rounds of entire genome duplications. However, Spring's hypothesis, later referred to as the 'one-to-four rule' (Ohno, 1999), or the '2R' hypothesis (Hughes, 1999) remains controversial (Hughes, 1999; Wang and Gu, 2000; Hughes *et al.*, 2001; other chapters in this issue).

Recently, an additional genome duplication event has been proposed in ray-finned fishes (Amores *et al.*, 1998; Wittbrodt *et al.*, 1998). The first indications for a fish-specific genome duplication came from studies based on *Hox* genes and *Hox* clusters. *Hox* genes encode DNA-binding proteins that specify cell fate along the anterior-posterior axis of bilaterian animal embryos and occur in one or more clusters of up to 13 genes per cluster (Gehring, 1998). It is thought that the ancestral *Hox* gene cluster arose from a single gene by a number of tandem duplications. Protostome invertebrates and the deuterostome cephalochordate *Amphioxus* possess a single *Hox* cluster, whereas Sarcopterygia, a monophyletic group including lobe-

finned fish, such as the coelacanth and lungfishes, amphibians, reptiles, birds, and mammals, have four clusters (Holland and Garcia-Fernandez, 1996; Holland, 1997). This finding has been regarded as important support for the '2R' hypothesis of two rounds of entire genome duplications early in vertebrate evolution. Recently, extra *Hox* gene clusters discovered in the zebrafish (*Danio rerio*), medaka (*Oryzias latipes*), the African cichlid (*Oreochromis niloticus*) and the pufferfish (*Takifugu rubripes*) suggest an additional genome duplication in ray-finned fishes (Actinopterygii) before the divergence of most teleost species (Amores *et al*., 1998; Wittbrodt *et al*., 1998; Meyer and Schartl, 1999; Naruse *et al*., 2000; Málaga-Trillo and Meyer, 2001; A. Amores, personal communication). In the meantime, comparative genomic studies have turned up many more genes and gene clusters for which two copies exist in fishes but only one copy in other vertebrates (e.g., Postlethwait *et al*., 2000; Robinson-Rechavi *et al*., 2001; Taylor *et al*., 2001a; Van de Peer *et al*., 2001; Woods *et al*., 2001). The observations that different paralogous pairs originate at about the same time (Taylor *et al*., 2001a), that they are found on different linkage groups, and that they show synteny with other duplicated genes (Gates *et al*., 1999; Postlethwait *et al*., 2000; Woods *et al*., 2000) support the hypothesis that these genes arose through a complete genome duplication event (Fig. 1a). On the other hand, several well-supported trees show one of the fish genes as the sister sequence to a monophyletic clade that included the second fish gene and genes from frog, chicken, mouse, and human (Taylor *et al*., 2001a; Robinson-Rechavi *et al*., 2001a). These so-called 'outgroup' topologies (Fig. 1d) might suggest that the origin of many fish duplicates predates the divergence of the Sarcopterygii and Actinopterygii and that tetrapods lost duplicates retained in fish (Fig. 1c).

Robinson-Rechavi *et al*. (2001a, 2001b) argued that an ancestral whole-genome duplication event was not responsible for the abundance of duplicated fish genes. They counted orthologous genes in fish and mouse and, where extra genes were found in fish, compared the number of gene duplications occurring in a single fish lineage with that shared by more than one lineage. Most mouse genes surveyed were also found as single copies in fish. Duplicated fish genes were detected, but most were interpreted as the products of lineage-specific duplication events in fish and not as an ancient duplication event (Fig. 1b). Here, we provide further evidence for the ancient fish-specific genome duplication based on phylogenetic inference, including sequences from multiple fish lineages.

Material and methods

Sequence alignments

Homologous sequences were collected and aligned as described before (Taylor *et al*., 2001a). In short, protein sequences were collected using BLASTp (Altschul *et al*., 1997) and aligned with CLUSTALX (Thompson *et al*., 1997). Sequence alignments were edited with BioEdit (Hall, 1999) and only unambiguously aligned regions were retained for further analysis. For this study, our aim was to collect homologous sets of genes that contained sequences from at least two different fish species. In most cases, genes from either zebrafish or pufferfish (*Takifugu rubripes*) were collected. Sequence alignments and additional data and information on sequence retrieval and analysis can be found in the Wanda database on duplicated fish genes (Van de Peer *et al*., 2002a; http://www.evolutionsbiologie.uni-konstanz.de/Wanda/).

Phylogenetic tree construction

In general, phylogenetic trees were constructed by neighbor-joining (Saitou and Nei, 1987) based on Poisson-corrected distances, as implemented in TREECON (Van de Peer and De Wachter, 1994). Recently, we developed a software tool called ASaturA to detect and consider saturation in amino acid sequences (Van de Peer *et al*., 2002b). When saturation is observed, evolutionary distances between sequences can be computed from the fraction of unsaturated sites only and evolutionary trees inferred by pairwise distance methods (for details, see Van de Peer *et al*., 2002b).

Results and discussion

Tree topologies support an ancient fish-specific genome duplication

Previously, we have shown that third-codon positions were saturated for most zebrafish paralogs (Taylor *et al*., 2001a; Van de Peer *et al*., 2001). Together with the observation that duplicated genes were found on different linkage groups, the most parsimonious

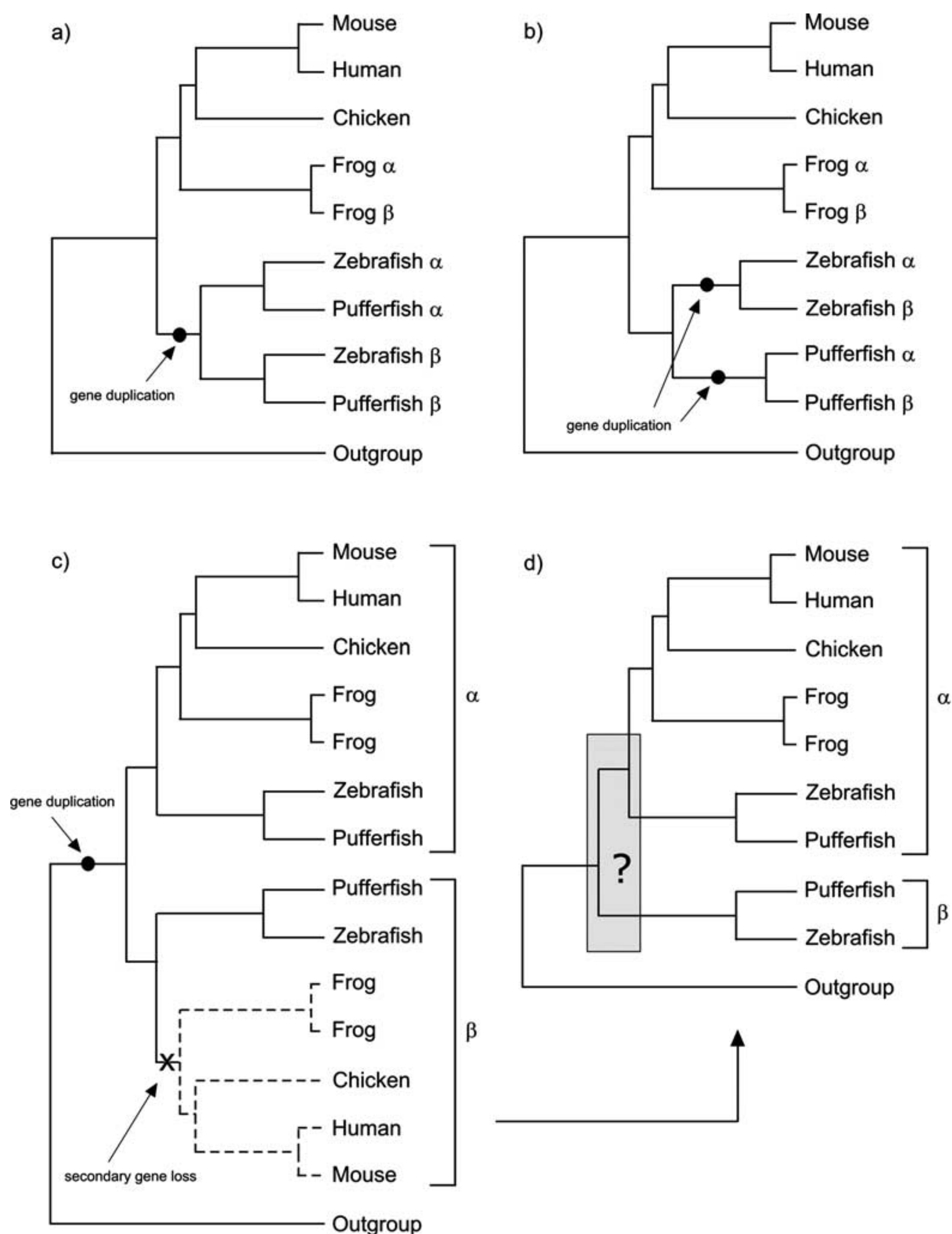

Figure 1. Different scenarios – and expected inferred tree topologies - to explain the presence of more genes in fish. (a) Duplicated fish genes are the result of a gene/genome duplication that preceded the divergence of zebrafish and pufferfish. (b) Duplicated genes are formed by independent gene duplications. The topology shown in d) is expected to be inferred when genes produced during a duplication event in the ancestor of Actinopterygii plus Sarcopterygii (shown in c) have been secondarily lost in the sarcopterygian lineage after the split of these two major lineages of jawed vertebrates. On the other hand, the tree topology shown in (d) might be an artifact in tree construction due to differences in evolutionary rates in the different duplicates (Taylor and Brinkmann, 2001) or due to saturation (Van de Peer *et al*., 2002b), which often makes it difficult to infer the exact branching order for deeper regions in the tree as indicated by the gray border.

explanation is that all these 'old' paralogs originated by an ancient genome duplication, somewhere between 300 and 450 million years ago (Taylor *et al.*, 2001a). Because major teleost lineages are believed to have arisen between approximately 100 and 200 million years ago (Carroll, 1997; Lydeard and Roe, 1997), the working hypothesis was to assume that the genome duplication occurred in the ancestor of most (if not all) ray-finned fish. To find additional evidence, we compiled many vertebrate data sets, including the zebrafish genes described previously (Taylor *et al.*, 2001a) and, when available, their pufferfish homologs. The almost complete pufferfish genome sequence has been made available recently (http:// www.jgi.doe.gov/) and pufferfish orthologs could indeed be found for most of the zebrafish genes. Zebrafish and pufferfish both belong to the Euteleostei (a Subdivision of the Superorder Teleostei), together with at least 22,000 others species and are rather distantly related among Euteleost fish. Zebrafish and pufferfish are estimated to have diverged approximately 150 million years ago (Nelson, 1994).

Figure 2 shows some of the inferred tree topologies, including duplicated genes from both zebrafish and pufferfish. In general, the trees shown (and deposited in the Wanda database) are Poisson-corrected distance trees taking into account all sites of the alignment. However, in some cases, improved tree topologies with higher statistical support could be obtained by removing saturated sites from the sequence alignment. Saturation was detected with ASaturA, a software tool specifically developed for this purpose. AsaturA is a Java-based application that visualizes the amount of saturation in amino acid sequences by graphically displaying the number of observed frequent and rare amino acid replacements between pairs of sequences against their overall evolutionary distance. Discrimination between frequent and rare amino acid replacements is based on substitution probability matrices (e.g., PAM and BLOSUM). When amino acid sequences showed saturation for a fraction of the sites, evolutionary distances were computed from the fraction of unsaturated sites only (for details, see Van de Peer *et al.*, 2002b).

Figure 2a shows a tree for *Reggie*, a cell surface protein found in retinal ganglion cells during axon regeneration. For the *Reggie* gene, also two paralogs from the goldfish *Carassius auratus* (Schulte *et al.*, 1997) were included. In Figure 2b a tree topology is presented for *DLL1*, a homolog of the distal-less gene in *Drosophila*, which is the first genetic signal for limb formation to occur in a developing zygote. The tree topology for *FZD8*, a family of putative transmembrane receptors homologous to the product of the *Drosophila* tissue polarity gene *frizzled*, and that inferred from a sequence alignment of *RXRB*, the retinoid X receptor β gene, are seen in Figures 2c and 2d, respectively.

As can be observed, all the tree topologies shown, taking into account either all sites (*DLL1* and *RXRB*) or only unsaturated fractions of sites (*Reggie* and *FZD8*) are in perfect agreement with an ancient fish-specific genome duplication that occurred before the divergence of zebrafish and pufferfish (see also Fig. 1a). Many additional trees, including duplicated genes from zebrafish and pufferfish with similar topologies can be found in the Wanda database (Van de Peer *et al.*, 2002a).

It should also be noted that, in order to test whether two species experienced the same gene or genome duplication, it is not necessary to find two genes in both species. If a gene from one species clusters specifically with one of the two duplicates (i.e. paralogs) of a second species, then this can only be explained by a shared duplication event with a subsequent loss of one of the gene copies in one of the species. Examples are given for the *SHH* gene (Fig. 2e), which codes for a signal that is necessary in patterning the early embryo, and for the *BMP2* gene, a highly conserved member of the transforming growth factor β gene family (Fig. 2f). For instance, the *BMP2* gene of *Takifugu rubripes* clusters specifically with one of the *BMP2* paralogs of *Danio rerio*. The second *Takifugu BMP2* gene is probably waiting to be discovered or has been secondarily lost.

The evolutionary consequences of an ancient fish-specific genome duplication: Gene duplication, functional divergence of genes, and speciation

If a fish-specific genome duplication had occurred, fish genomes would be expected to contain more genes, at least initially, than the genomes of mammals. In our genome survey, we very often uncovered multiple gene copies in fish for single genes in other vertebrates, but almost never the opposite (see also Wittbrodt *et al.*, 1997: Robinson-Rechavi *et al.*, 2001a, 2001b). After at least 200 or more million years of evolution, these duplicated fish genes might be expected to have acquired quite different functions. Ohno's model, which Hughes (1994) first

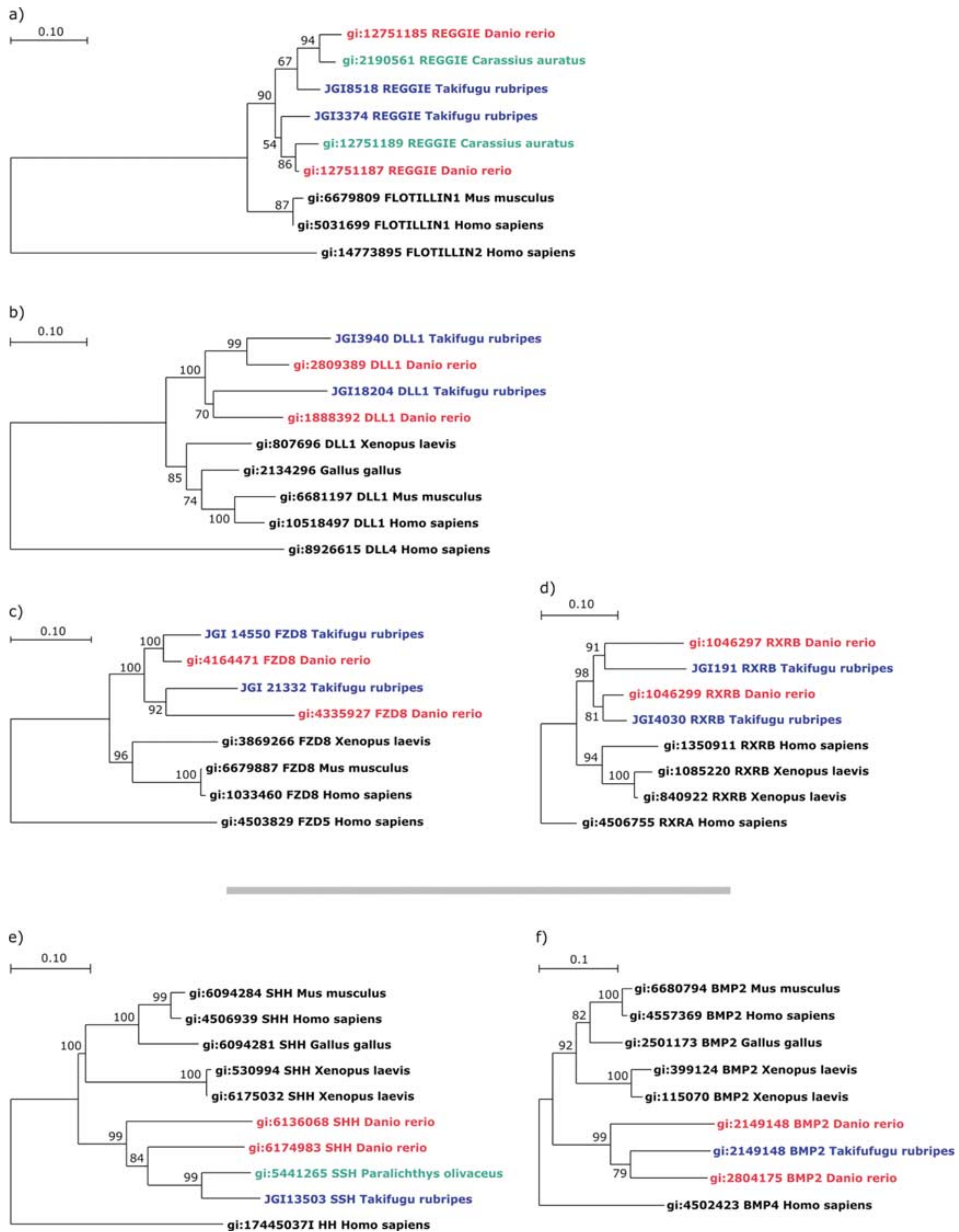

Figure 2. Selected phylogenies including genes of multiple fish lineages. All tree topologies, including either two or more pairs of duplicated fish genes (topologies a-d) or one pair of duplicated genes plus (a) single additional gene(s) from other fish species (topologies e-f) support an ancient fish-specific genome duplication. Additional trees supporting the fish-specific genome duplication can be found in the Wanda database (Van de Peer *et al*., 2002a).

called the 'mutation during non-functionality' and later the 'mutation during redundancy' model (Hughes, 1999), predicts that, if genes do not get lost, by chance, a series of non-deleterious mutations might render the duplicate gene into a gene with a new function. This model has been widely adopted to explain the evolution of functionally novel genes, but little evidence can be found that new gene functions have evolved this way. Although it might seem unlikely that anciently duplicated genes perform completely redundant functions, redundancy has been shown to be widespread in genomes of higher organisms (Nowak *et al.*, 1997 and references therein; Gibson and Spring, 1998). Furthermore, many paralogous fish genes seem to have subdivided their functions rather than to have evolved novel functions. Recently, the 'duplication-degeneration-complementation' (DDC) model (Force *et al.*, 1999; Lynch and Force, 2000a) has been introduced to explain why duplicated genes might be retained. This model predicts that the likelihood of preservation is correlated with the number of 'subfunctions' that can be ascribed to a gene. The model starts from the assumption that a gene can perform several different functions, for instance, expression in different tissues and at different times during development, each of which may be controlled by different DNA regulatory elements. If duplicate genes lose different regulatory subfunctions, each affecting different spatial and/or temporal expression patterns, then they must complement each other by jointly retaining the full set of subfunctions that were present in the ancestral gene. Therefore, degenerative mutations facilitate the retention of duplicate functional genes, where both duplicates now perform different but necessary subfunctions. However, as predicted by the DDC model, the sum of the retained duplicates must be equal to the total number of subfunctions performed by the ancestral gene. Gene duplication then allows each daughter gene to specialize for one of the functions of the ancestral genes. Force *et al.* (1999) showed that this model might generally apply based on the *En1* genes in zebrafish. In mouse and chicken, *En1* is expressed in the developing pectoral appendage bud and in specific neurons of the hindbrain and spinal cord (Joyner and Martin, 1987; Davis *et al.*, 1991; Gardner and Barald, 1992). In zebrafish, however, one of the paralogs is expressed in the pectoral appendage bud, while the second paralog is expressed in the hindbrain/spinal cord neurons (Force *et al.*, 1999).

Possibly, retention of gene duplicates by subfunctionalization applies to many of the anciently duplicated fish genes. Besides *En1*, differences in the expression pattern of *Msx* zebrafish paralogs and homologous genes of other vertebrates also suggest subfunctionalization of the zebrafish genes after duplication (Ekker *et al.*, 1997). Similar conclusions can be drawn for *hedgehog* genes (Laforest *et al.*, 1998), *Bmp2* (Martinez-Barbera *et al.*, 1997), the transcription factors *mitfa* and *mitfb* (Mellgren and Johnson, 2002; Altschmied et al., 2002), *cyp19* (Chiang *et al.*, 2001), *GlyRalpha* genes (Imboden *et al.*, 2001), *Notch* and *Pax6* (Lynch and Force, 2000a). Models such as the DDC model may explain the retention and functional divergence of duplicated genes. However, when paralogs diverge in function mainly through subfunctionalization, functional divergence is probably limited to differences in timing and tissue specificity of expression. Until now, there is little evidence that the fish paralogs have changed functions completely in the course of evolution. Therefore, it is still an open question whether subfunctionalization of many duplicated genes resulting from the fish-specific genome duplication can be responsible for the large number of fish species and their tremendous morphological diversity, as suggested previously (Amores *et al.*, 1998; Wittbrodt *et al.*, 1998; Meyer and Schartl, 1999).

However, another phenomenon could explain the abundance of fish species we observe. Recently, a model called 'divergent resolution' has been proposed (Lynch and Conery, 2000; Lynch and Force, 2000a), in which the loss or silencing of duplicated genes might be more important to the evolution of species diversity than the evolution of new functions in duplicated genes. Divergent resolution occurs when different copies of a duplicated gene are lost in geographically separated populations and could genetically isolate these populations, should they become reunited (reviewed in Taylor *et al.*, 2001b; Fig. 3). Therefore, large-scale gene duplications and rapid speciation of organisms might be correlated. In this respect, it is noteworthy that also in plant evolution there is a strong indication for a polyploidy event that seems to coincide with a massive diversification of novel plant families (Raes *et al.*, this issue; Y. Van de Peer, unpublished data).

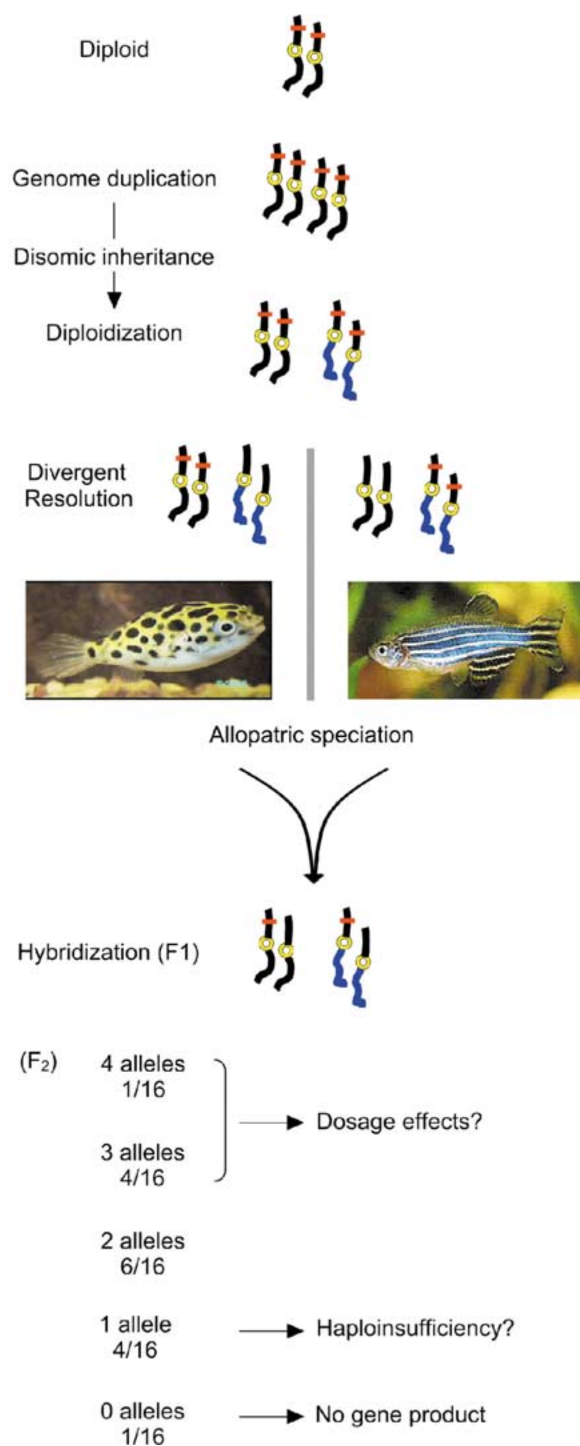

Figure 3. Speciation through genome duplication and divergent resolution. Natural selection will favour speciation over hybridization in populations fixed for different copies of a duplicated locus. Red bars represent a locus that is duplicated (along with all other loci) during a tetraploidy event. In this hypothetical example, diploidization is driven by a reciprocal translocation depicted by a change in chromatid colour. Pufferfish (left) and zebrafish (right) are shown as examples of the descendants of the two populations. If individuals from such populations mate, their 'hybrid' progeny would be heterozygous, possessing a functional allele and a pseudogene at each locus of the duplicated gene. Crosses between the F1 individuals produce some (about 6%) F2 individuals with only pseudogenes at both loci in question, and therefore lacking viability and/or fertility. Others would receive between one allele, which might lead to reduced function when the gene product from one functional allele is inadequate to support normal function (haploinsufficiency), to three or four functional alleles, which might have a negative dosage effect. All these might lead to postmating reproductive isolation (Lynch and Force, 2000b). Reprinted (and slightly adapted) from Trends in Genetics 17, pp. 299–301, © 2000, 'Genome duplication, divergent resolution, and speciation' by John S. Taylor, Yves Van de Peer, and Axel Meyer, with permission from Elsevier Science.

Acknowledgements

This work was supported by the German Science Foundation (DFG PE 842/2–1). J.S.T. is indebted to the National Sciences and Engineering Research Council of Canada for a Postdoctoral Fellowship. A.M. received support from the Deutsche Forschungsgemeinschaft (DFG), the University of Konstanz and the Fonds der Chemischen Industrie.

References

Altschmied, J., Delfgaauw, J., Wilde, B., Duschl, J., Bouneau, L., Volff, J.-N. and Schartl, M. (2002) Subfunctionalization of duplicate *mitf* genes associated with differential degeneration of alternative exons in fish. *Genetics* **161**, 259–267.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*., **25,** 3389–3402.

Amores, A., Force, A., Yan, Y.-L., Joly, L., Amemiya, C., Fritz, A., Ho, R.K., Langeland, J., Prince, V., Wang, Y.-L., Westerfield, M., Ekker, M. and Postlethwait, J.H. (1998) Zebrafish *hox* clusters and vertebrate genome evolution. *Science*, **282,** 1711–1714.

Aparicio, S., Hawker, K., Cottage, A., Mikawa, Y., Zuo, L., Venkatesh, B., Chen, E., Krumlauf, R. and Brenner, S. (1997) Organization of the *Fugu rubripes Hox* clusters: evidence for continuing evolution of vertebrate *Hox* complexes. *Nature Genet*., **16,** 79–83.

Carroll, R.L. (1997) *Patterns and Processes of Vertebrate Evolution*, Cambridge University Press, Cambridge, UK.

Chiang, E.F., Yan, Y.L., Tong, S.K., Hsiao, P.H., Guiguen, Y., Postlethwaith, J. and Chung, B.C. (2001) Characterization of duplicated zebrafish *cyp19* genes. *J. Exp. Zool*., **290,** 709–714.

Davis, C.A., Homyard, D.P., Millen, K.J. and Joyner, A.L. (1991) Examining pattern formation in mouse, chicken and frog embryos with an *En*-specific antiserum. *Development*, **2,** 287–298.

Ekker, M., Akimenko, M.A., Allende, M.L., Smith, R., Drouin, G., Langille, R.M., Weinberg, E.S. and Westerfield, M. (1997) Relationships among *msx* gene structure and function in zebrafish and other vertebrates. *Mol. Biol. Evol*., **14,** 1008–1022.

Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.-l. and Postlethwait, J. (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics*, **151,** 1531–1545.

Gardner, C.A. and Barald, K.F. (1992) Expression patterns of engrailed-like proteins in the chick embryo. *Dev. Dyn*., **193,** 370–388.

Gates, M.A., Kim, L., Cardozo, T., Sirotkin, H.I., Dougan, S.T., Lashkari, D., Abagyan, R., Schier, A.F. and Talbot, W.S. (1999) A genetic linkage map for zebrafish: comparative analysis and localization of genes and expressed sequences. *Genome Res*., **9,** 334–347.

Gehring, W.J. (1998). *Master Control Genes in Development and Evolution: the Homeobox Story*. Yale University Press, New Haven

Gibson, T.J. and Spring, J. (1998) Genetic redundancy in vertebrates: polyploidy and persistence of genes encoding multidomain proteins. *Trends Genet*., **14,** 46–49.

Hall, T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser*., **41,** 95–98.

Holland, P.W. (1997) Vertebrate evolution: something fishy about *Hox* genes. *Curr. Biol*., **7,** R570-R572.

Holland, P.W.H. (1999) The effect of gene duplication on homology. In *Homology* (Eds., Bock, G.R. and Cardew, G.), Wiley, Chichester, UK, pp. 226–242.

Holland, P.W. and Garcia-Fernandez, J. (1996) *Hox* genes and chordate evolution. *Dev. Biol*., **173,** 382–395.

Hughes, A.L. (1994) The evolution of functionally novel proteins after gene duplication. *Proc. R. Soc. Lond. B*, **256,** 119–124.

Hughes, A.L. (1999) Phylogenies of developmentally important proteins do not support the hypothesis of two rounds of genome duplication early in vertebrate history. *J. Mol. Evol*., **48,** 565–576.

Hughes, A.L., da Silva, J. and Friedman, R. (2001) Ancient genome duplications did not structure the human Hox-bearing chromosomes. *Genome Res*., **11,** 771–780.

Imboden, M., Devignot, V. and Goblet, C. (2001) Phylogenetic relationships and chromosomal location of five distinct glycine receptor subunit genes in the teleost *Danio rerio*. *Dev. Genes Evol*., **211,** 415–422.

Joyner, A.L. and Martin, G.R. (1987) *En–1* and *En–2*, two mouse genes with sequence homolog to the *Drosophila engrailed* gene: expression during embryogenesis. *Genes Dev*., **1,** 29–38.

Laforest, L., Brown, C.W., Poleo, G., Geraudie, J., Tada, M., Ekker, M. and Akimenko, M.-A. (1998) Involvement of the *Sonic Hedgehog*, *patched* 1 and *bmp2* genes in patterning of the zebrafish dermal fin rays. *Development*, **125**, 4175–4184.

Lundin, L.-G. (1999) Gene duplications in early metazoan evolution. *Cell Dev. Biol*., **10,** 523–530.

Lydeard, C. and Roe, K.J. (1997) The phylogenetic utility of the mitochondrial cytochrome b gene for inferring relationships among actinopterygian fishes. In *Molecular Systematics of Fishes* (Eds., Kocher, T.C. and Stepien, C.A.), Academic Press, San Diego, CA, pp. 285–303.

Lynch, M. and Conery, J.S. (2000) The evolutionary fate and consequences of duplicate genes. *Science*, **290,** 1151–1155.

Lynch, M. and Force, A. (2000a) The probability of duplicate gene preservation by subfunctionalization. *Genetics*, **154,** 459–473.

Lynch, M. and Force, A. (2000b) The origin of interspecific genomic incompatibility via gene duplication. *Am. Nat*. **156**, 590–605.

Málaga-Trillo, E. and Meyer, A. (2001) Genome duplications and accelerated evolution of *Hox* genes and cluster architecture in teleost fishes. *Amer. Zool*., **41**: 676–686.

Martinez-Barbera, J.P., Toresson, H., Da Rocha, S. and Krauss, S. (1997) Cloning and expression of three members of the zebrafish Bmp family: *Bmp2a*, *Bmp2b* and *Bmp4*. *Gene,* **198**, 53–59.

Mellgren E.M. and Johnson, S.L. (2002) The evolution of morphological complexity in zebrafish stripes. *Trends Genet*., **18,** 128–134.

Meyer, A. and Schartl, M. (1999) Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. *Curr. Opin. Cell Biol*., **11,** 699–704.

Naruse, K., Fukamachi, S., Mitani, H., Kondo, M., Matsuoka, T., Kondo, S., Hanamura, N., Morita, Y., Hasegawa, K., Nishigaki, R., Shimada, A., Wada, H., Kusakabe, T., Suzuki, N., Kinoshita, M., Kanamori, A., Terado, T., Kimura, H., Nonaka, M. and Shima, A. (2000) A detailed linkage map of medaka, *Oryzias latipes*: comparative genomics and genome evolution. *Genetics*, **154,** 1773–1784.

Nelson, J.S. (1994) *Fishes of the World*, 3rd ed., Wiley, New York, NY.

Nowak, M.A., Boerlijst, M.C., Cooke, J. and Maynard Smith, J. (1997) Evolution of genetic redundancy. *Nature*, **388,** 167–171.

Ohno, S. (1970) *Evolution by Gene Duplication*, Springer Verlag, New York, NY.

Ohno, S. (1999) The one-to-four rule and paralogues of sex-determining genes. *Cell. Mol. Life Sci*., **55,** 824–830.

Patel, N.H. and Prince, V.E. (2000) Beyond the *Hox* complex. *Genome Biol*., **1,** 1027.1–1027.4.

Postlethwait, J.H., Woods, I.G., Ngo-Hazelett, P., Yan, Y.-L., Kelly, P.D., Chu, F., Huang, H., Hill-Force, A. and Talbot, W.S. (2000) Zebrafish comparative genomics and the origins of vertebrate chromosomes. *Genome Res*., **10,** 1890–1902.

Robinson-Rechavi, M., Marchand, O., Escriva, H., Bardet, P.-L., Zelus, D., Hughes, S. and Laudet, V. (2001a) Euteleost fish genomes are characterized by expansion of gene families. *Genome Res*., **11,** 781–788.

Robinson-Rechavi, M., Marchand, O., Escriva, H. and Laudet, V. (2001b) An ancestral whole-genome duplication may not have been responsible for the abundance of duplicated fish genes. *Curr. Biol*., **11,** R458-R459.

Saitou, N. and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol*., **4,** 406–425.

Schulte, T., Paschke, K.A., Laessing, U., Lottspeich, F. and Stuermer, C.A. (1997) Reggie–1 and reggie–2, two cell surface proteins expressed by retinal ganglion cells during axon regeneration. *Development*, **124,** 577–587.

Sidow, A. (1996) Gen(om)e duplications in the evolution of early vertebrates. *Curr. Opin. Genet. Dev*., **6,** 715–722

Spring, J. (1997) Vertebrate evolution by interspecific hybridisation — are we polyploid? *FEBS Lett*., **400,** 2–8.

Taylor, J.S. and Brinkmann, H. (2001) 2R or not 2R. *Trends Genet*., **17,** 488–489.

Taylor, J.S., Van de Peer, Y., Braasch, I. and Meyer, A. (2001a) Comparative genomics provides evidence for an ancient genome duplication in fish. *Phil. Trans. Roy. Soc. B*, **356,** 1661–1679.

Taylor, J.S., Van de Peer, Y. and Meyer, A. (2001b) Genome duplication, divergent resolution, and speciation. *Trends Genet*., **17,** 299–301.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res*., **25,** 4876–4882.

Van de Peer, Y., and De Wachter, R. (1994) TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. *Comput. Appl. Biosci*., **10,** 569–570.

Van de Peer, Y., Taylor, J.S., Braasch, I. and Meyer, A. (2001). The ghost of selection past: rates of evolution and functional divergence in anciently duplicated genes. *J. Mol. Evol*., **53,** 434–444.

Van de Peer, Y., Taylor, J.S., Joseph, J. and Meyer, A. (2002a) Wanda: A database of duplicated fish genes. *Nucleic Acids Res*., **30,** 109–112.

Van de Peer, Y., Frickey, T., Taylor, J.S. and Meyer, A. (2002b) Dealing with saturation at the amino acid level: A case study based on anciently duplicated zebrafish genes. *Gene*, **295**, 205–211.

Wang, Y. and Gu, X. (2000) Evolution patterns of gene families generated in the early stage of vertebrates. *J. Mol. Evol*., **51,** 88–96.

Wittbrodt, J., Meyer, A. and Schartl, M. (1998) More genes in fish? *BioEssays*, **20,** 511–512.

Woods, I.G., Kelly, P.D., Chu, F., Ngo-Hazelett, P., Yan, Y.-L., Huang, H., Postlethwait, J.H. and Talbot, W.S. (2000) A comparative map of the zebrafish genome. *Genome Res*., **10,** 1903–1914.